

Energy-efficiency versus resilience: risk awareness view on dimensioning of optical networks with a sleep mode

Piotr Cholda¹ · Piotr Jaglarz¹

Received: 28 July 2014 / Accepted: 14 March 2015 / Published online: 11 April 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract This article proposes to solve the trade-off between energy-efficiency and resilience with a focus on business mechanisms. Risk engineering is used as a foundation. Financial impact (penalty) quantification with various compensation policies is applied, and business-relevant risk measures are used during the risk assessment. Then, risk mitigation strategies are evaluated to select the appropriate risk response. The approach is presented in networks with energy profiles supporting a sleep mode. An effective heuristic is used to assign flows, and it is shown that the energy-efficiency performance is substantially independent of the recovery methods selected for risk mitigation. It is also demonstrated that backup resources can be switched off in the normal state without having a considerable impact from a financial viewpoint.

Keywords Energy-efficiency · Green networks · Network recovery · Reliability · Risk engineering

1 Introduction

Failures result in either node or link outages. In effect, some services are interrupted, which in turn is responsible for financial losses for operators. To counter these losses, automatic *recovery methods* are designed. They operate in various ways (with different levels of sharing, scope, etc. [20]) and result in different parameters, such as *resilience* (i.e.,

survivability to random failures) related to client needs. They also incur increased capital expenditure (CAPEX) and operational expenditure (OPEX) costs, since in order to bypass fault-affected elements, it is necessary to use *backup* resources. Here, we focus on one of the most prominent OPEX costs: the cost of energy usage. Idzikowski et al. [34] rate resilience (and reliability-related parameters) among the main criteria of assessing energy-aware design. The reason is that large optical pipes carry a lot of traffic which should not be left interrupted without a response.

In this paper, we deal with dimensioning of energy-efficient resilient backbone optical networks. Such networks aim to minimize energy usage or prevent the consumption of non-renewable energy (e.g., produced by traditional coal plants). From the business viewpoint, there are at least three aspects encouraging energy-efficiency [2]: (a) high costs of energy; (b) if energy usage is too high, its supply can simply be cut off, which can replace capacity as a bottleneck in network management and operation [10]; (c) pressures on the industry to protect the environment may force regulators to introduce energy-saving policies. All these aspects are important for the operators rather than the clients, who simply require uninterrupted services. Wiatr et al. [64] elaborate on the existence of the trade-off between energy-efficiency and quality performance, demonstrating that power minimization cannot always be the main driver in network design. Likewise, resilience involves a trade-off, since its introduction applies increased energy usage due to the introduction of backup resources. Here, we propose to approach this trade-off with the application of some elements of *risk engineering* [21] and to look at this problem from the business perspective. Therefore, we shift the focus from reducing energy usage (while satisfying various demands) to assessing the potential impact of *making networks environmentally friendly (green)* on the resilience perceived by a network

This work was funded by the NCBiR and CHIST-ERA ERA-Net SwiTching And tRansmission project.

✉ Piotr Cholda
piotr.cholda@agh.edu.pl

¹ AGH University of Science and Technology, Kraków, Poland

client. Then, we are able to deal with the mitigation of recognized risks (failures) combined with energy-efficiency. We are not aware of any other approach of combining the problems of energy-efficient recovery methods with their selection based on risk awareness.

Precisely, we deal with network dimensioning problems (in uncapacitated networks) using Yaged's heuristic approach providing fast solutions that can be obtained due to the assumed energy profiles supporting *sleep modes*, typical for elements of optical networks. Our optimization bears in mind that during a normal, failure-free network state (more than 99 % of the time), there is no need to consume energy. Thus, by shifting the interest from capacity (no longer important due to the pervasive overprovisioning of optical resources) to the energy usage viewpoint, we are able to show that the most attractive options for the abovementioned trade-off solution are quite surprising. Dedicated protection provides the most benefits; however, using a sleep mode with spare (backup) resources with re-routing appears to be an interesting option.

The remainder of the paper is organized as follows. First, to show the context of our results, we discuss related work in Sect. 2. Section 3 presents energy profiles that are assumed in our optimization and performance evaluation studies. The optimization for network dimensioning is based on a fast design method and provides energy-efficient routing with various recovery methods. Section 4 outlines risk management: the framework we use to deal with the intrinsic trade-off between energy-efficiency and resilience provisioning (selected risk engineering aspects). After presenting its organization (risk management cycle), we focus on various methods of expressing the monetary impact of failures (compensation policies) and methods for meaningful quantification of the predicted financial losses (risk measures), as well as business-relevant selection of countermeasures (risk mitigation). Section 5 shows the impact of the selected risk mitigation strategies on various recovery settings. As the results assume one recovery method for all demands, in Sect. 6 we propose a simple optimization method to find a combination of recovery methods for the assumed mitigation strategy. In Sect. 7, we formulate general conclusions on how the assignment will look like and should be carried out. Finally, we provide a summary with a view on future work in the closing Sect. 8.

2 Related work

2.1 Energy-efficiency in resilient networks

While the first papers on greening (improving environmental credentials) of networks appeared in the early 2000s, here we limit our presentation to relatively recent relevant literature. The field of network energy-efficiency is described in many

excellent surveys [7, 9, 49, 67]. They present the justification for dealing with energy-efficiency in optical backbone networks. They give details on the main methods to provide it and present different approaches.

Sleep mode is one of the methods of consolidating resources [7], where traffic is consolidated in some network elements while the other elements are put in sleep mode (i.e., some links are made unavailable). Wiatr et al. [63] discuss various types of sleep (in passive optical networks): power shedding, deep/fast sleep, and dozing, with all their advantages and disadvantages. However, we deal with the problem from a more abstract perspective. Network sleep mode is typically assumed with the design of energy-aware networks, where the diminishing return effect (also known as bandwidth discount [12]) typically takes place. It is suggested that the mode is used in a twofold manner: (a) on a short time scale (traffic engineering), where it is adjusted to short-term load variability (e.g., over a day), and (b) on a longer time scale (network dimensioning), where it is used for spare resources to be switched on when they are needed after failures occur. The latter approach is relevant to this paper. While generally the sleep mode mechanisms are not present in contemporary network devices, some attempts to include them in protocol suites exist. For instance, Morea et al. [44] propose enhancements in the Generalized Multi-Protocol Label Switching (GMPLS) control plane protocols. Additionally, documents have been produced in the Internet Engineering Task Force by Eman (a working group on energy management, established in 2009) to provide the management plane with the necessary ontology to represent data useful in energy-aware networks, taking into account the sleep mode [22, 52]. Energy profiles supporting our approach and adapted in the design and optimization of optical networks, sometimes with resilience provisioning, are reported in [5, 9, 13, 14, 36, 44, 56, 67]. While assuming the most common fixed + proportional ($f + p$) energy profile for single elements, the total energy usage for the whole network has the concave character as a function of the summarized network load [36, 67]. An in-depth study of energy usage is also given in [59], where a very simple rule of thumb for energy usage is proposed as a universal value for the electrical layer (1 Gb/s costs 10 W for a MPLS router); in the optical layer, it is necessary to consider optical switching and calculate the number of transponders and regenerators on links. It is shown that power efficiency is obtained by reducing marginal energy usage per unit capacity of a card port, thus justifying using of concave energy profiles. Chiaraviglio et al. [14] consider some risks related to using the sleep mode based on the $f + p$ energy profile. Using elegant theoretical modeling, they show that sleep is advantageous unless the fixed cost of switching on components is an order of magnitude lower than the proportional energy cost.

It should also be noted that sometimes reported energy profiles are different than those relevant to this work. For instance, Kilper et al. [37] take into account fixed switching on costs, but they also assume that the unit cost of variable energy usage increases with capacity; as such they consider convex characteristics, at least in access networks. They claim that for capacities lower than 10 Gb/s, the power of line cards is not dependent on capacity, but that for larger capacities, there is some dependency. Convexity also concerns the electronic level: there is a square or cubic relationship with the frequency or voltage [8, 28, 53]. However, at the network element level, it seems that the effect is reduced to $f+p$.

Restrepo et al. [53] analyze various energy profiles: on-off (step function), logarithmic-like functions (representants of concave functions whose usage justifies the application of the sleep mode), cubic-like (convex) characteristics responsible for representing increases in energy usage with voltage or frequency in electronic circuits, and purely linear functions (some switching devices are said to have this property). They use all these energy profiles in the minimized goal functions and present how to route traffic accordingly. They present linear optimization problems applying a very coarse linearization (with two segments only) in order to not increase the complexity of the linear problem too much. To some extent, we refine their approach with a more sophisticated optimization of concave functions and additionally propose an efficient heuristic to obtain solutions of a high quality (a very small optimization gap).

While literature regarding resilience and energy-efficiency issues separately is extensive, papers discussing both together also exist. However, they typically deal with optimization problems from the energy-efficiency minimization viewpoint, considering resilience needs as additional constraints [47]. From this perspective, it has been shown that energy-efficiency approaches (mainly using sleep modes for spare resources) enable operators to save a significant percentage of energy usage, even with protection methods applied, but the reliability performance is put to jeopardy with the increased sleep rate. While it has also been noted that energy minimization and resilience provisioning are opposing goals and some trade-offs are needed (e.g., by using money saved by the decreased energy usage to finance the increased costs of recovery as proposed by Wiatr et al. [62]), no proposals for countering this have been provided. We attempt to fill this gap by using the risk-aware approach by mapping both energy and reliability performance into one monetary scale. Many papers provide optimization formulations with $f+p$ energy profiles; however, to our knowledge, none of them propose the use of Yaged's heuristic. A typical approach in the context of resilient energy-efficient networks is to minimize energy usage with mixed integer linear programming (MILP) formulations such as in [44], where routing and wavelength assignment (RWA) is opti-

mized jointly with 1:1 dedicated protection, and then the optimization formulation is solved by a commercial solver. For instance, Liu et al. [38] present algorithms for energy-efficient routing of connections protected with shared backup paths, so that the shared spare resources are put to sleep. They also take into account shared-risk link groups (SRLG), a notion important in multilayer networks, not to be confused with our risk management approach (two links are elements of the same SRLG if they fail together). Lopez et al. [39] work on the fact that provisioning various service classes to clients provides some gains to operators (e.g., in energy savings or capacity costs). They focus on energy usage and deal with various recovery classes. Nevertheless, the authors use fixed scenarios with given percentages of clients assigned various service classes. In contrast, we consider how to provide a given client, defining their needs with a specific compensation policy, with a recovery method conforming to the operator's risk mitigation strategy. Muhammad et al. [45] investigate switching off devices that are used for recovery purposes only. We follow this approach in relation to re-routing, but in the case of protection methods, we assume that sleeping elements cannot be woken on demand. Francois et al. [29] propose a new method that provides resilience while also aiming to use energy more efficiently. They propose the use of green backup paths to carry some traffic (so that other links can go sleep); they are used again as real backup paths when a failure happens (this resembles a preemptive mode in 1:1 automatic protection switching standardized for transport networks). They also present a two-goal optimization problem (by minimizing the maximum link usage and maximizing the total amount of energy saved), as well as solving the RWA problem. The fact that energy-efficiency as each cost-focused goal must be traded off with desirable non-functional properties has been noted before. For instance, Cavdar et al. [13] pay attention to the resilience-energy trade-off by noting that the use of sleep modes and the focus on reducing the number of elements used increase the number of single points of failure. Additionally, as traffic carried via some elements is greater than if the sleep mode is not applied, an average failure affects the operation in a more significant way. Jirattigalachote et al. [36] consider a dedicated path protection 1:1, where links carrying backup resources are switched off while they are not used. To ensure that it is possible to put a considerable number of links to sleep, they aim to separate links with working and backup resources. They note the trade-off between energy minimization and performance (the probability of blocking) in their dynamic case. Wiatr et al. [64] note the fact that energy-aware network design mainly focuses on power minimization, while other aspects should be traded off with energy considerations. They also focus on quality issues, such as request blocking. Finally, Vizcaino et al. [61] also note the trade-off between energy-efficiency and resilience constraints (represented as service

availability). They analyze various approaches to the design of path protection mechanisms (both dedicated and shared) in elastic optical networks. Similar to [50], they skip re-routing due to its long operation following a failure. It is worth noting that the papers recognizing the trade-offs do not consider the financial aspect of transferring the opposing aspects into monetary units to express the goal functions in a unified way and thus supporting the decision.

2.2 Risk awareness in resilient networks design

Franke [30] notes that generally the discussion of the relationship between the technical aspect and the business context related to the management of telecommunication networks is poorly developed. While the methods and protocols for network resilience, described for instance in [20,55,60], are not a new topic, a business-oriented approach to resilient network design has not been studied in depth, and many problems remain. Risk had originally been dealt with in two types of industries: financial (where investment portfolios are being selected) and technological (where failures jeopardize human safety or societal welfare). The two approaches have been combined in the telecommunications sector (selection of new investments [26], or in security against faults generated by malicious behavior [1]). We focus on resilience provisioning, a topic covered in general in [21]. However, in the context of risk management in resilient networks, risk assessment is the most popular topic. While the value-at-risk measure was postulated to be used in communications networks for resilience quantification [1,21,43], this has little reference to optimization methods used in the financial sector, where value-at-risk is basic [41].

A proposal of linear optimization of risk mitigation strategies is outlined in [18]. Vajanapoom and Tipper [58] present a set of papers on the topic, with the most comprehensive presenting linear programming-based models applying the risk exposure measure. A similar average-based risk measure in network design is proposed in [48]. Dikbiyik et al. [25] present an integer linear program that also bases risk response on risk exposure, where the consequences are based on the cumulative downtime exceeding the assumed threshold. Gonzalez and Helvik [32] provide a set of optimization approaches using two-stage stochastic programs to increase the provider's gain (to minimize recovered connection costs and penalties paid for outages incurred). As mentioned before, these approaches are typically not based on an exact analysis of quantile risk measures relevant to business. An approach somewhat similar to ours is given in [24], which considers large-scale failures (disasters) in cloud environments. It presents various approaches to optimized design of disaster-resilient cloud infrastructures, when one of the approaches is based on risk minimization.

To the authors' knowledge, the only work that combines energy-efficiency problems with optimization methods specific for risk management has been presented by Cano et al. [11]. However, the approach is totally different to ours. Apart from the fact that the authors deal with energy-efficient planning of heating/cooling systems in buildings rather than with the networking problems, risk management is related to the investment process (the risks are related to a lack of energy sources, etc.) and not to adverse events such as failures. In contrast to our approach, they develop a two-state stochastic optimization problem usually encountered in risk-based investment planning, for instance in the chemical industry [6,68].

3 Design of resilient energy-efficient networks

We take a typical assumption that energy can be saved by using a *sleep mode*. There are three options for the operation of devices with respect to power management procedures investigated by researchers [50]: (a) active mode (fully used): at each time point, the device can operate at full capacity; (b) sleep mode (low power/standby/idle/hibernation): the device is using some energy and can switch to active mode almost instantaneously; (c) switched off mode (inactive): the device is not using any energy, and it takes some considerable time for the device to be woken up. While the difference between the last two modes is important in traffic engineering, here we are interested in network dimensioning (long-term behavior), and we identify the sleep mode with switching off [36]. Perello et al. [50] deal with two basic types of sleep modes: link sleeping mode (LSM) and optoelectronic device sleep mode (OESM). The latter assumes switching off transponders or regenerators, while the former mainly concerns optical amplifiers. Since the switching on process in OESM can be fast (order of milliseconds), it may be suitable for protections. LSM (with wake-up time taking seconds) is too slow for protection methods, and we take it into account for re-routing methods only. Later, we study the influence of this assumption on risk assessment. Another aspect relating the sleep mode to resilience is stressed by Caria et al. [12], who emphasize that due to switching off, traffic paths are longer, increasing delays, susceptibility to failures, and degrading connectivity.

From the optimization viewpoint, the sleep mode is feasible only with selected *energy profiles*, i.e., energy usage characteristics or functions of traffic load being carried through a component. Ricciardi et al. [54] partition energy profiles into three groups: (a) experimental: presenting real-world data taken from device vendors (declarations) or operators (measurements), etc.; (b) analytical: models focused on an operating environment, omitting many details and configurations while still presenting exact characteristics; and (c)

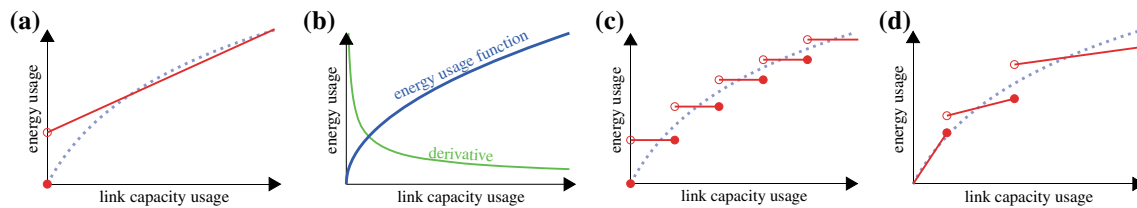


Fig. 1 Various energy profiles supporting sleep modes from the optimization viewpoint. *Dotted curves* represent the concave approximations. **a** Fixed+proportional ($f+p$) energy profile (with an approximating concave curve). **b** Concave energy profile with its derivative,

illustrating the *decreasing return to scale*. **c** Energy profile for ALR: step function (with an approximating concave curve). **d** Energy profile for ALR: linear stepwise function or piecewise concave cost (with an approximating curve)

theoretical: the simplest abstractions that grasp the most general behavior of the devices. Here, we use the last type, since we want to show a new methodology of heuristic-based optimization and the risk mitigation-based approach. We posit that dealing with more complex energy profiles may hinder the presentation and unnecessarily lengthen the paper, and the influence on the general results will be low. While it is assumed that the desirable energy profile should be linear (i.e., with a zero cost at origin and the slope proportional to the load), in practice other cases are encountered that ensure the feasibility of the sleep mode. For almost all devices [49], energy profiles consist of the static (fixed) part, active only when a link is used, and the dynamic part dependent on various parameters (frequency, voltage, or traffic load). The simplest model adequate for the sleep mode assumes that when the link is not used, the energy usage EU is zero; then, if a link is used, the fixed cost E_0 is significant and the dynamic part is proportional with a constant E_p to the link load L . We call this model $f+p$, i.e., fixed+proportional. The mathematical definition of $f+p$ is as follows (see Fig. 1a):

$$EU = \begin{cases} 0 & \text{if } L = 0 \\ E_0 + E_p \times L & \text{if } L > 0 \end{cases} \quad (1)$$

Generally, each model with a fixed cost or *concave* dynamic part favors the sleep mode, since in this case, it is energy-efficient to increase the loads in some elements and switch off others (by decreasing traffic flows carried via them to zero). A concave function is characterized by decreasing marginal energy usage per traffic capacity unit, being responsible for the economical effect known as the *decreasing return to scale* (see Fig. 1b). Here, we assume that the energy profile of all links in networks we study is simply an increasing concave function, since it not only approximates $f+p$ well, but also is justified for many other important cases. This type of energy usage is reported in some transport technologies (as mentioned in Sect. 2), and it also approximates the behavior of energy profiles characteristic for adaptive/multiple line rate (ALR) approaches [37]. The latter are proposed as Ethernet 802.11az, or they are relevant for optical networks

with wavelength division multiplexing (WDM), where some wavelengths and related transponders can be switched off; additionally the wavelengths themselves can have various rates. The relevant models are illustrated in Fig. 1c–d.

We assume the static network dimensioning scenario in a transparent optical network, where for a given network we look for routing lightpaths for a long horizon, and where the traffic demand matrix does not change. The aim is to minimize total energy usage. We do not want to focus on RWA aspects; therefore, we assume that all flows can attain continuous values, and routing can be described with linear constraints. Nevertheless, as the goal is to minimize energy usage constructed out of concave energy profiles, the optimization problem is known as a difficult *concave optimization* task. In such a case, it can be linearized by providing additional constraints and binary variables that are used to model the concave function with linear segments. The problem can be represented in the form of an *NP-hard* MILP optimization task. The general methodology is described in [51, ch. 5]. The inclusion of many binary variables is prohibitive for major problems, since it is not possible to obtain the exact solution in practice. Therefore, to perform the numerical studies, we decided to use a modified Yaged's method [51, ch. 5.6], a heuristic polynomial algorithm with rapid convergence, designed to solve problems where the concave goal function is minimized. The method is outlined in Fig. 2. The algorithm allocates flows iteratively using shortest paths. We use Dijkstra's algorithm for single paths and Suurballe's algorithm for the shortest cycles (necessary to find backup paths for protections). Here, "shortest" means "minimizing energy usage", although for comparison we also show results based on distance and hop-count minimization. Yaged's method is based on the observation that concave energy costs mean that it is better to use loaded links rather than opening new ones. As a result, the demand flows are not bifurcated, i.e., they are routed with single paths, a feature highly desirable in optical networks. The complete solution provides a set of links that do not carry traffic (they can be switched off).

Each link e (in a network consisting of E links), carrying traffic load y_e , has its own concave energy profile

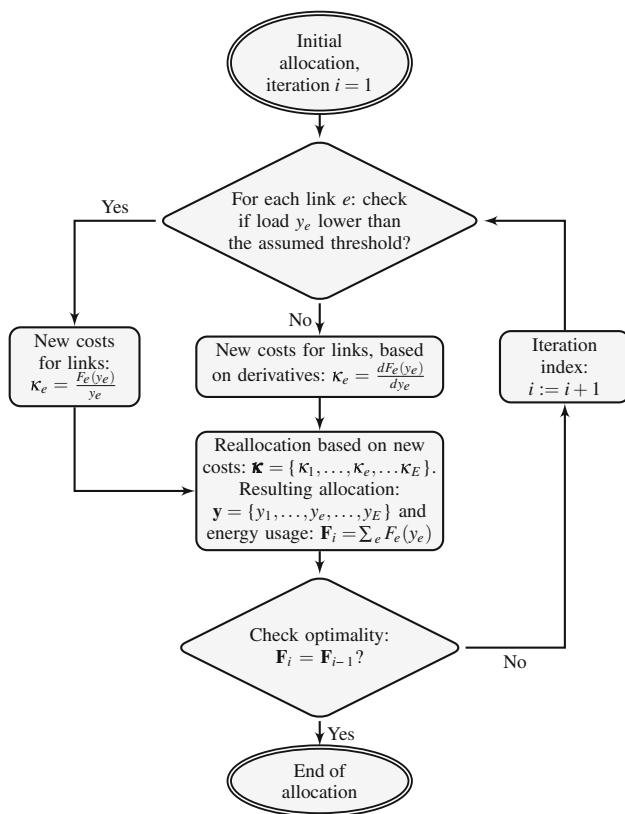


Fig. 2 A sketch of routing optimization with a modified Yaged's method

$F_e(y_e)$. First, the algorithm calculates the initial flow allocation $\mathbf{y}^* = \{y_1^*, \dots, y_e^*, \dots, y_E^*\}$ (e.g., simply with the shortest hop routing), and then the output energy usage: $\mathbf{F}_1 = \sum_e F_e(y_e^*)$. Then, the weights of links κ are calculated as derivatives of energy usage found for the current link loads: $\kappa_e = \frac{dF_e(y_e)}{dy_e}|_{y_e=y_e^*}$. The new output energy usage found on the basis of shortest routing with the set of link weights κ , i.e., $\mathbf{F}_2 = \sum_e F_e(y_e^*)$ will be lower than \mathbf{F}_1 . The algorithm iterates this way until energy usage ceases to decrease. To make the solutions independent of the initial allocation, we use a modified algorithm, where to calculate a new weight for a link with a low traffic load (below an assumed threshold), we do not use the derivative, but rather the tangent of the secant between the origin and point $(y_e^*, F_e(y_e^*))$: $\kappa_e = \frac{F_e(y_e^*)}{y_e^*}$.

Yaged's method performs well not only for concave cost functions, but for all cost functions that have non-increasing derivatives of increasing costs. Therefore, it can be used not only for simple concave approximates of the $f+p$ profile, but directly for such a profile. This profile has a derivative that is constant (E_p) except for the Dirac delta function in 0, which can be modeled as a rapidly decreasing monotonous function dependent on the relation between E_0 and E_p . The character of this decrease is not a problem from the perspective of calculation complexity, as application of Yaged's method performs a simple numerical calculation where the

derivative can be given as a set of numbers, and symbolic calculations are not necessary. We apply Yaged's method to the $f+p$ profile in Sect. 5.

4 Risk management in resilient network design

Network designers are experienced in recovery methods and mathematical tools to describe and optimize their operations. However, this specialist work is influenced by business management decisions on how money should be spent. The interface between the technological and business levels is handled by *risk management*, providing design results useful for the main business, i.e., ensuring the continuity of the operator's mission in the presence of adverse events. One of the relevant aspects is to match recovery methods to client needs defined in the service level agreement (SLA). From this perspective, it may be unreasonable not to provide recovery to clients whose SLAs predict a high level of penalties if the connections are interrupted. Similarly, it does not make sense to provide 1 + 1 protection (the most reliable option from the engineering viewpoint) to all clients, since the fees charged to many clients do not balance the cost for such protection, making it an unnecessary cost for operators [19,39].

Dealing with client needs is discussed during the *risk management cycle*. It follows the plan–do–check–act philosophy of the Deming cycle [18,65] and organizes how businesses approach the analysis and optimization of responses (recovery methods in our case) to adverse random events (failures), which are described by risk engineering. Risk is understood as “an uncertain event or condition that, if it occurs, has a positive or negative effect on an objective” [35], where each operator's objective is to increase profits and minimize losses. As both are dependent on random events, the risk is described with two basic parameters: by *probability* and *impact*, the latter best expressed in monetary (financial) units. Therefore, the first stage of the risk management cycle is *risk assessment* to identify types of risks (e.g., node/link failures) and evaluate their parameters (i.e., using statistical methods) with the aim of expressing the impact as penalties imposed on the operator with the relevant measures. After recognizing and prioritizing risks, it is possible to control them using *risk response*, first choosing feasible methods (e.g., preselecting specific recovery methods relevant to a given technology) and then deciding how to combine them to conform to policies assumed by the network operator while assessing the predicted decrease in risk and costs involved. Finally, the countermeasures should be *deployed* (involving the re-configuration of resources, testing, etc.) and *monitored* (to check whether the response implemented meets the intended goals). These two closing stages are less interesting from the network dimensioning viewpoint; therefore, we do not discuss them.

4.1 Risk assessment

Energy is one of the costs, and its minimization is the concern of the operator, while high levels of resilience are in the client's interest. During risk management, it is possible to trade off these opposing aspects. If the client is risk-neutral [31], the best way is to express them with the same monetary measure. If there is a known energy profile, the energy cost can be easily expressed monetarily, and the way to express the monetary impact of not conforming to the client's needs is to use penalties. A model that transforms technical loss (as perceived by the client) into its monetary equivalent is defined by *compensation policies* [43]. A penalty is paid to the clients affected by failures if such a case is predicted in their SLA. The most typical compensation policy is to assume that the penalty is proportional to the cumulative downtime over a given interval. As this policy can be expressed by *interval availability*, we call it Δ_V . An opposite approach is to base the policy on the number of all outages perceptible at the service level over a given interval. This means that the interest is shifted to *continuity* rather than availability; this applies to highly demanding services which are rendered useless by a failure no matter how fast the recovery method (a case for real-time control [15]). We call this policy Co . These two extreme policies can be combined. For instance, it can be assumed that the penalty is based on the number of outages exceeding a selected downtime threshold [25] or that the penalty is not scaled proportionally over an interval [23]. As this is the first approach to energy-efficient design with risk awareness, we stick to the basic policies in order not to overcomplicate the view.

Failure is a random event; therefore, the value of a penalty is a random variable. Risk measures are the means for predicting the behavior of network connections before and after deploying the selected risk response. Various known risk measures can be chosen [21,66]. The simplest to interpret is the mean penalty per interval, known as *risk exposure* (RE): it is the average amount of penalties that reduce the profit. It is also *coherent* [4], a useful feature in the context of convex optimization [17]. Nevertheless, it does not grasp the variability of the impact or extreme values (i.e., two totally different distributions may have the same mean value [31]). Hence, quantile risk measures have been proposed. The most fundamental is value-at-risk (VaR), the maximum penalty with a given confidence interval. VaR is commonly accepted in investment management for which many optimization methods elaborated in *modern portfolio theory* exist. They have also been proposed for networking [1,21,33,43]. Let ξ be the level of penalties, pertaining to a single connection or a whole network. If $P_\xi(x) = \Pr\{\xi \leq x\}$ is the cumulative distribution function of ξ , VaR is defined as the maximum penalty with a given confidence interval η : $VaR_\eta = P_\xi^{-1}(\eta) = \inf\{x : \Pr\{\xi \leq x\} \geq \eta\}$. Although VaR

is theoretically not a coherent risk measure [1,4,21], another study [17] shows that from a practical perspective, it behaves as a coherent measure. While the value of RE can be treated as a mean measure of operator's profit reduction over a given interval, the value of a quantile risk measure can be treated as a suggestion how much money should be buffered in the worst case to deal with the results of adverse events over an interval. The latter approach is typically assumed in the risk-based project management [27].

4.2 Risk response

An *informed business decision* on risk response is to invest in recovery if the penalties are high or recovery costs are low. From the business viewpoint, basic responses (e.g., risk avoidance, transfer, or acceptance [16,18]) can be very different, but a crucial option in the design of resilient networks is *risk mitigation*. It involves recovery methods to decrease the risk R by reducing impact (i.e., penalties) of failures. However, other than from the operator's viewpoint, both the levels of R and B (the energy budget/cost for risk mitigation/recovery provisioning) are taken into account as an important goal. These two objectives are contrary to each other, and there are two basic methods to deal with such cases [70]: either (a) constructing an aggregate objective function (with weights assigned to various goals) or (b) modeling one objective as a constraint in the optimization of another. Risk mitigation uses the first approach by weighing both goals with monetary values. We follow this path since energy profiles and compensation policies enable us to define monetary equivalents for both of them. This approach is also more flexible, since in the risk-neutral contexts, it enables us to relax the requirements to some extent. The operator still needs to pay penalties, but it is not necessary to define restrictive thresholds on resilience that cannot be exceeded.

The aggregation approach finds a convex characteristic, representing decreasing Pareto-optimal (non-dominated) solutions. Formally, option (r_1, b_1) , with risk level r_1 and related budget b_1 , dominates the option (r_2, b_2) if and only if: $(r_1, b_1) \succeq (r_2, b_2) \Leftrightarrow r_1 < r_2 \wedge b_1 \leq b_2$. The Pareto front, i.e., a theoretical curve presenting non-dominated options, is shown in the budget-risk (B, R) space in Fig. 3. The level of risk before deploying the response is known as *baseline risk* R_{base} , while the level below which it is not possible to drop (independently of the budget, since failures will happen) is known as *residual risk* R_{res} . The following *business strategies* to select risk mitigation are elaborated in the IT security field (they are summarized in Table 1). (a) *Risk minimization*, RM [3]: the aim is to downsize risk as much as possible ($\min R$) for the lowest possible cost: $\text{argmin}_B \left(\frac{dR(B)}{dB} = 0 \right)$. From the theoretical viewpoint, if treated as absolute minimization, this option is perceived as extremely costly and

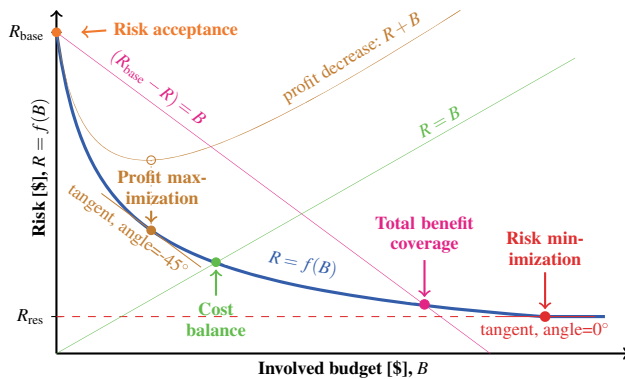


Fig. 3 Basic risk mitigation strategies shown in the (B, R) space. The blue line marks the non-dominated solutions. Line $R = B$ is the oblique asymptote for the profit decrease curve

not useful in a typical carrier's practice (except for operating critical infrastructures). However, in Sect. 5, we show that this is not true for the specific networks with concave energy costs. (b) *Total benefit coverage*, TC [3]: the cost of organizing the strategy is equal to the risk reduction: $D = R_{\text{base}} - R(B)$. This means we look for the budget so that $\text{argmin}_{B \neq 0} |D = B|$, and as a result, uncertain risk is exchanged for a known (certain) cost. This can also be presented with linear constraints using properties of non-dominated solutions, since we are interested only in values of D better than the budget B (which means the solution will be feasible if the curve is below the $(R_{\text{base}} - R) = B$ line in Fig. 3, which is not always the case), and we select the option with the highest D level. (c) *Cost balance*, CB [57]: the aim is to find a point where the risk level and the budget involved are the same: $|R(B) = B|$. This is recommended as a strategy for finding business continuity for US federal institutions; again, we can present this strategy with linear constraints ensuring that the risk is greater than the budget $B \leq R$ and finding the minimum risk level. (d) *Profit maximization*, PM [3]: the aim is to maximize the total monetary gain. Looking for such a solution is a well-known economical problem of *net utility maximization* [40], where utility is the risk reduction and the price is the budget involved. Mathematically, the solution is the point where the marginal risk reduction is balanced by the marginal budget increase or $\min(R(B) + B)$. We also show *risk acceptance*, RA, which is not a risk mitigation strategy; however, we use it as a benchmark, since it is a budget optimization solution.

5 Numerical example, part 1: coarse-grained optimization

First, we optimize the use of resources from the viewpoint of energy usage. We demonstrate that the modified Yaged's heuristic performs well in comparison with the exact

Table 1 Summary of the risk mitigation optimization approaches

Strategy	Additional constraints	Goal function
Risk acceptance	(none)	$\min B$
Risk minimization	(none)	$\min R$
Total benefit coverage	$B \leq D = R_{\text{base}} - R$	$\max D$
Cost balance	$B \leq R$	$\min R$
Profit maximization	(none)	$\min (R + B)$

optimization. Second, we show simulation results of real energy usage and risk values and look for solutions for the assumed risk mitigation strategies. While these results are being discussed for protections, we also check how re-routing with switching off backup resources performs from the risk awareness viewpoint. It appears that re-routing outperforms protection methods if only the amount of backup resources in switched on links exceeds a specific threshold.

The network and connections within it are modeled with a very simple approach, i.e., a network is represented by a graph with a set of perfectly reliable nodes and a set of unreliable links connecting the nodes. A demand is defined as a requirement to carry a volume of data between two nodes. It should be fully satisfied with a single non-bifurcated connection. Network topologies and demands are retrieved from the SNDlib library (<http://sndlib.zib.de>) as models of two different networks: PL, the Poland Network (polska.xml) representing a sparse topology; and Ge, the German Research Network (nobel-germany.xml) representing a dense topology.¹ Each demand is assigned a single recovery option out of the following set. (a) *No recovery* (NR): all demands are carried in a way optimizing energy usage, which is treated as basic cost B_0 ; if failures affect connections, they are interrupted. (b) *Dedicated path protection 1:1* (DP): the connections are routed with two disjoint paths, where the whole set of connections optimizes energy usage; in the normal state, the shorter of the disjoint paths is used as a working path, and the other (a backup path) is not used. This means that although the backup capacity must be reserved, it does not use energy if not needed [36]. (c) *Dedicated link protection 1:1* (DL): the working paths are routed as for NR, but backup segments for protecting working capacity in links are added (also assuming energy usage optimization). (d) *Shared backup path protection* (SP): a pair of paths for the connections is found as for DP; however, energy usage optimization takes into account that backup resources are shared among working paths. To calculate the usage, a single failure assumption is taken; this is challenged in simulations

¹ Due to a lack of space, we are not able to present all the results here; therefore, others are given here: <http://home.agh.edu.pl/~cholda/research/energy-efficiency-vs-resilience/>. The results obtained for PL and Ge networks are similar in nature; therefore, we do not show those relating to Ge in the paper.

Table 2 Comparison of exact versus heuristic optimization for the PL network

Recovery option	Goal: CPLEX [C]	Goal: modified Yaged's heuristic [Y]	Relative difference: $\frac{Y-C}{\min\{Y,C\}}$ (%)	Number of iterations of the Yaged's heuristic
NR	528.04	522.06	−1.15	6
DP	971.77	972.09	0.03	4
DL	1186.40	1174.04	−1.05	2
SP	820.62	801.50	−2.38	3
SL	900.62	936.37	3.97	1

where multiple failures can be present. (e) *Shared backup link protection* (SL): it differs from DL in that backup capacity is shared among segments protecting the working capacity of various links (again, an absence of multiple failures is assumed during optimization). (f) *Re-routing (restoration) with $x\%$ backup capacity reserved in working links and an option to switch on sleeping links*: working paths are routed as in NR, then the links not being used are put to sleep; they can be woken up if it is not possible to re-route the affected connections using the backup capacity reserved in switched on links (the amount of the reserved backup capacity equals to $x\%$ of the working capacity in a link). As for protections, the re-routing is studied in path (RP_x) and link (RL_x) versions. We do not present results for dedicated path/link protections 1 + 1, set analogously to DP/DL, but rather with traffic steadily repeated to backup resources (also in the normal state). From the risk perspective, these methods offer little improvement (with the existing technological solutions, the switching time is negligible); therefore, they are clearly dominated as being the most costly. Other methods can be applied in practice [20]; however, here we focus only on those that can be modeled in a compact way, and we show that even with this small set of options, non-trivial results can be obtained.

We assume that the energy profile is modeled with the concave *square root function* of a load on a link. The profile is linearized with linear segments, where each has the length of two in abscissa (representing the traffic load in a link). Table 2 compares results of the optimization of energy usage when calculated with the exact MILP formulation with the optimization software CPLEX 12.5 (on a server with an Intel Xeon E5-2680@2.93 MHz processor, 24 GB of RAM, and 12 cores) with those obtained by the MATLAB application of the Yaged's method. It was not possible to solve the exact model in any case (memory overflow due to many branching processes); therefore, a level of relaxation gap for branch and bound was assumed to not exceed 10% and set prior to conducting the calculations. The optimization process took several hours (typically around 24h), but it was able to provide results either up to 4% better than with the heuristic Yaged's method or even worse. It has been demonstrated that Yaged's method converges after a finite number of steps; in our case, this is fewer than ten, and the whole process takes no more than 0.6 s.

Next, MATLAB was used to develop simulations for finding dynamic energy usage and risk values, assuming that the paths are assigned as those found during the optimization stage. Two distributions are associated with each link: the first describes exponential time between failures, while downtimes are modeled with the Pareto distribution. The distribution parameters are proportional to link lengths [60], and their values are retrieved from [42]. The independent failure and repair interchanging process is simulated in each link, and then the effect on connections (with the assumed recovery option in operation) is calculated with both compensation policies. Each connection has its own parameters necessary to find the exact value of the penalty, which is assumed to be the product of downtime hours (Δv) or number of outages (C_o) and the demand volume given with network models (in Mb/s). When risk values are calculated for the whole network, the distributions of the penalties estimated for individual connections are summed. For each simulation scenario, we assume that all the connections use the same recovery option; we held 1000 simulations over 100,000 h, enabling us to find the distributions of penalties and the risk measures (RE and $VaR_{0.95}$). The penalty distributions do not have heavy tails, a phenomenon generally found in resilient networks [17]. For each simulation time point, we calculate the link loads and transform them with assumed square root energy profiles to the temporary energy usage. The risk measures and mean energy usage are expressed for an interval of 10 years.

A difficult problem is finding meaningful monetary coefficients to obtain a financial value of a unit of energy cost and penalty. Penalty coefficients depend on the market, SLAs in place, and technical network conditions (topologies, failure rates, etc). For the authors, as researchers working at academia, the challenge of accessing such data has been too great. Therefore, we use values found online. Taking into account that the presented estimation may be inexact, we also study results obtained when the orders of these coefficients are modified. The base values of coefficients are obtained as follows. We assume that an approximate cost of one kilowatt-hour in the USA is 0.2 USD,² and following [46], we take

² Electric power monthly: http://www.eia.gov/electricity/monthly/epm_table_grapher.cfm?t=epmt_5_6_a.

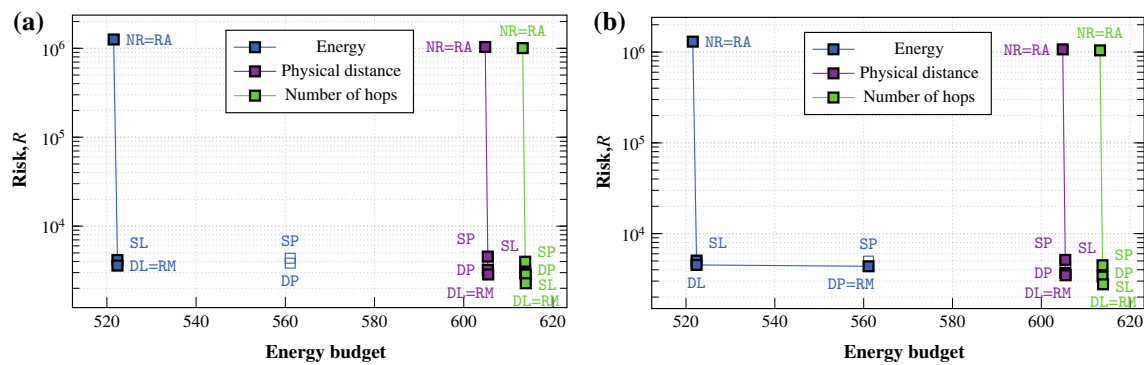


Fig. 4 Summary of simulation results of various protection methods (the A_v compensation policy is used to find the penalties) in the PL network with concave energy profiles. *Filled squares* denote the non-dominated solutions (they are joined by *lines*), while the *empty squares*

represent dominated solutions. The risk/energy coefficients do not affect domination. In all the cases: $RM = PM = CB = TC$. **a** Risk measure: RE . **b** Risk measure: VaR

the approximate energy for sending 1 Gb/s as being around 100 W on average, taking into account all the optical components. Carrying all the demands with NR generates a load of 26,243 Mb/s. During 10 years of operation (approximately 100,000 h), the total energy cost is 0.27119 million USD, which gives the energy unit coefficient of 520 USD, since optimization results give a value of 521.5 energy units. We also assume that the rental cost of 100 Mb/s optical fiber is approximately 1500 USD per month,³ which gives a total cost per hour of work of 1 Gb/s as approximately 20.5 USD. According to the given penalty policy,⁴ we assume that the penalty for 1 h of downtime is equal to the reduction in the payments of one day, which means the risk coefficient for A_v is 24 times the cost of the rental per hour that is approximately 500 USD. To find the risk coefficient for C_o , we apply the rule presented in [43], where the reimbursement for a single outage is the product of the reimbursement for a unit of downtime multiplied by the average duration of an outage. In the case of NR, the latter is 10h, which means that the risk coefficient for C_o is 5000 USD. That is, we assume that approximately the risk to energy unit coefficient is equal to 1 in the case of A_v , and 10 in the case of C_o .

Simulation results related to protections are shown in Fig. 4 (note that the abscissa gives the total energy used, i.e., the involved risk mitigation budget can be found by subtracting the amount of energy necessary to provide RA). The first surprising result is that for various risk measures, the relative position of solutions to various strategies differs very little, an effect quite opposite to the results obtained for the budgets based on capacity reserved reported in [16]. As for

protections we assume to use OESM that switches on very fast, we do not penalize the cases when it is necessary to increase link capacities, and in this case, A_v is a more adequate compensation policy. The results basing budgets on energy usage appear to be almost invariant of the optimization method (note that energy-efficiency promotes much larger paths, which are more prone to failures), and some of the optimum points overlap. This stems from the fact that with concave energy profiles, it is easy to provide a risk minimization strategy (in all cases, it is DL) that surpasses other strategies, since the difference in risk of various recovery options (except for NR) is very low, and the most reliable one is preferred. Even if the initial optimization is conducted with a classic capacity-oriented approach (minimization of the summarized number of hops) or reliability-oriented approach (minimization of the summarized physical distances reducing failure rates), the results show almost no differences from the solution perspective, although the numbers differ. It should be noted that although optimization changes, mitigation is still based on energy usage. Another surprising fact is that in such a coarse-grained approach to risk mitigation (i.e., one recovery option for all), the theoretical characteristic shown in Fig. 3 is highly degenerated. We show in Sect. 7 that this effect is less pronounced if the optimization takes into account reserved capacity (i.e., an approach taken in classic network dimensioning) rather than energy usage. It is necessary to emphasize that the degeneration effect is not related to the used monetary coefficients, but to the character of the budget consumption, where in our case, the energy is consumed only when the links are used. Due to very large penalty in the NR case, the monetary coefficients do not have the impact on the solutions. It is also notable that although local protections are more costly during capacity usage optimization, they are less costly from the dynamic energy usage perspective than their path counterparts. Additionally, local versions provide a higher reliability (i.e., decrease the risk),

³ BTnet Leased Line: <http://business.bt.com/broadband-and-internet/leased-lines/>.

⁴ BTnet Service Level Agreement: http://business.bt.com/assets/pdf/broadband-and-internet/datasheet/BTnet_leased_line_service_level_agreement.pdf.

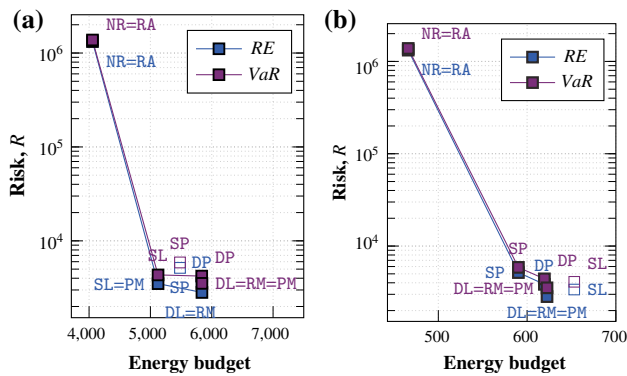


Fig. 5 Summary of simulation results of various protection methods (the A_v compensation policy is used to find the penalties) in the PL network with two $f+p$ energy profiles. *Filled squares* denote the non-dominated solutions, while the *empty squares* represent dominated solutions. In all the cases: $RM=CB=TC$. **a** OTN-100 Gb/s profile. **b** OTN-10 Gb/s profile

since they are more resilient to multiple failures. The path methods (DP and SP) use very long working and backup paths, since this is more advantageous from the energy usage perspective (more links are loaded); this effect also appears in the case of the backup segments protecting the links, although it is not as strong. Therefore, the link methods dominate.

As the use of the concave energy profile may seem controversial, we decided to provide the results also for a plain $f+p$ profile. We used two profiles of this kind. They are reported in [59]. The first energy profile (known as OTN-100 Gb/s) has a very big fixed cost ($E_0 = 360$ W) and a small proportional cost ($E_p = 3.6$ W/Gb/s). The second profile (known as OTN-10 Gb/s) has almost ten times smaller fixed cost ($E_0 = 34$ W), and a very similar proportional cost ($E_p = 3.4$ W/Gb/s). Some results related to the use of these profiles are shown in Fig. 5. We can see that the energy values are different (we do not rescale the energy coefficients), but the qualitative character of the results is analogous to ones presented in Fig. 4, although it is possible to see a modest influence of the risk measure on the final results (domination and a solution for the strictly business-related PM strategy). This stems from the fact that the effect of scale is again present and that the energy is consumed mostly in the failure-free situations. Therefore, we can assume that the presented results are valid for a very broad range of energy profiles supporting the sleep mode.

We assess protections separately from the re-routing options since their establishment and operation philosophies are different: for protections, we find paths that use the shortest cycles which makes the working paths longer and increases energy usage, while re-routing uses the shortest paths. This also involves various protocols, signaling procedures, and switching times. However, some general comparisons are also instructive. In the case of re-routing, we check the fact that they may use resources in switched

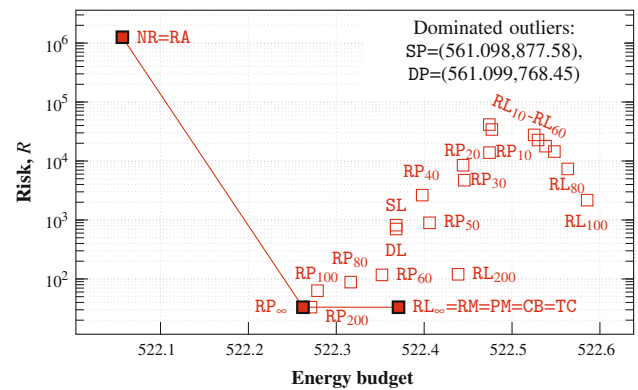


Fig. 6 Summary of simulation results of various re-routing and protection methods (the C_o compensation policy is used to find the penalties) in the PL network with concave energy profiles. Seven of 18 links are switched off in the normal state for re-routing methods. *Filled squares* denote the non-dominated solutions, while the *empty squares* represent dominated solutions. Routing is based on energy usage. The risk measure applied is RE

off links. Contrary to [13,36] which claim the usefulness of switching off backup resources, Perello et al. [50,61] state that switching off the backup resources fully is not acceptable from the reliability viewpoint. Here we investigate the real impact of the opposite assumption. We assume that switching on the sleep resources is triggered when backup resources are necessary in the failure state, since other backup resources are consumed in fully utilized links or there is no physical connectivity to redirect a connection for an affected demand. In our case, whole WDM links are put to sleep, and contrary to the protection scenario, we assume to use the LSM mode. Therefore, switching on a WDM link takes a lot of time and considerably interrupts the connection. We consider it as an outage and quantify the related penalty with the C_o compensation policy.

The results are shown in Fig. 6 (while the results are presented for RE , the character of results for VaR does not change). An interesting fact about the re-routing options RP_x/RL_x can be observed: the greater the capacity reserved as backup in switched on links (i.e., ones in a non-sleep mode in the normal state), the lower the probability that it will be necessary to switch on sleeping links and therefore the total energy usage is lower. Hence, from the energy viewpoint, a good option is to switch off as many links as possible and switch them on only if it is not possible to route the connections that should be restored with non-sleeping links. This effect is shown as decreasing energy usage with increasing backup rate x . The risk also decreases simultaneously, which is a phenomenon more obviously significant in the case of the C_o policy. In fact, the best option is to use RL_∞ which needs to wake up additional links only when there is no physical connectivity in the basic set of links due to multiple failures. Additionally, we can see that there is a threshold x (here,

$x = 60\%$) when the re-routing starts to dominate the protection methods and is most advantageous for an operator from the risk mitigation viewpoint. However, the latter finding is valid for long agreements, which may be contrary to the assumptions supporting the Co policy [15]. We can generalize these results to the two following extreme cases: if reservation of capacity is costly (e.g., it must be leased), then protection still makes much sense. If, on the other hand, it is possible to overprovision network capacity and the operator's cost is related mainly to the energy usage (e.g., since the network is owned by this operator), re-routing is the preferred solution. Again, the results stem from the scale effects due to concavity of energy profiles.

6 Risk mitigation: fine-grained optimization in energy-efficient networks

Here, we propose a simplified approach to finding the optimal combination of recovery options, assuming that each connection can use its own option, contrary to the coarse-grained approach presented so far. Since in the PL network, there are 66 connections (121 for the Ge network), the number of all possible combinations, where each connection may have a different recovery option (skipping re-routing), is 5^{66} . They cannot all be analyzed by simulations to obtain the exact risk and energy values. This is why we decided to base the calculations on linear programming with risk and energy values estimated in five simulation scenarios presented in Sect. 5, where all the connections apply a single recovery option. Output numbers will not be precise, since risk measures are nonadditive for quantile measures (even though they may be subadditive in practice [17]) and for shared protections, the resulting values are pessimistic (as the combination reduces the sharing of backup resources). Another simplification is that we assume linear aggregation of energy costs. We do not have an easy method for including concave energy costs in the constraints taking into account goal functions containing risk (unlikely to be the case when the concave energy usage function is the only element of the goal function as relevant in the linearization methods mentioned in Sect. 3). Although the approach is simplistic, a somewhat similar approach is successfully used in the project risk management field [69]. A more precise approach based on modeling of the total reserved capacity is outlined in [16]. Demands are denoted with d (there are D demand overall), and the recovery options used are denoted with t (so that $t \in \{\text{NR}, \text{DP}, \text{DL}, \text{SP}, \text{SL}\}$). The constant r_{dt} gives the value of the risk measure obtained for demand d if it applies recovery option t , while the constant e_{dt} gives the energy consumption of demand d when it uses recovery option t . The overall energy usage if all demands are assigned NR (risk acceptance) is denoted as B_0 , while the baseline risk in this case is R_{base} . All the constants are given

in monetary units (thus, we do not disturb the formulation with energy/risk monetary coefficients). The binary variable x_{dt} equals 1 if demand d is assigned recovery method t , and 0 otherwise. The constraints common to all the optimization programs to find solutions for different mitigation strategies are as follows:

$$\sum_t x_{dt} = 1 \quad d = 1, \dots, D \quad (2)$$

$$\sum_d \sum_t e_{dt} x_{dt} - B_0 = B \quad (3)$$

$$\sum_d \sum_t r_{dt} x_{dt} = R \quad (4)$$

$$R_{\text{base}} - R = D \quad (5)$$

Equation (2) enforces that each demand is assigned only one recovery method. Equation (3) defines the budget involved as the total cost of the enforcement of the combination of various recovery methods minus the cost of the risk acceptance case. Equation (4) determines the value of the risk incurred by the combination of various recovery options assigned to the connections, while Eq. (5) gives the value of the risk decrease. To finalize the formulation for an assumed mitigation strategy, it is necessary to add the goal function and additional constraints describing an assumed risk mitigation strategy, as enumerated in Table 1.

7 Numerical example, part 2: fine-grained optimization

CPLEX was used to develop optimization problems presented in the previous section to obtain solutions with mixed recovery options assigned to various demands. Although the problem assumes that MILP is used, the calculations are very fast (each takes less than 2s) as there are only up to a few hundred binary variables. On the basis of the five basic configurations, where all the demands are assigned only one recovery option (NR, DP, DL, SP, and SL), we find risk values r_{dt} for all connections and the relative input to the total energy usage, e_{dt} . The latter is calculated as a sum over all links used by a demand, and proportional to the total energy used in the link and the demand volume.

We assume that although the demands can use various recovery options, they all apply the same compensation policy and risk measure. In this case, the optimization results presented in Fig. 7 are based on the Av policy (since the combined recovery options are limited only to protections, and for them, we assume application of this policy since the recovery switching is fast) and RE (again, the results obtained with VaR are similar). We show the results for various values of the monetary coefficients by which we multiply a unit of energy used, capacity reserved, or penalty. The upper part shows the results of optimization if the budget is based on the energy

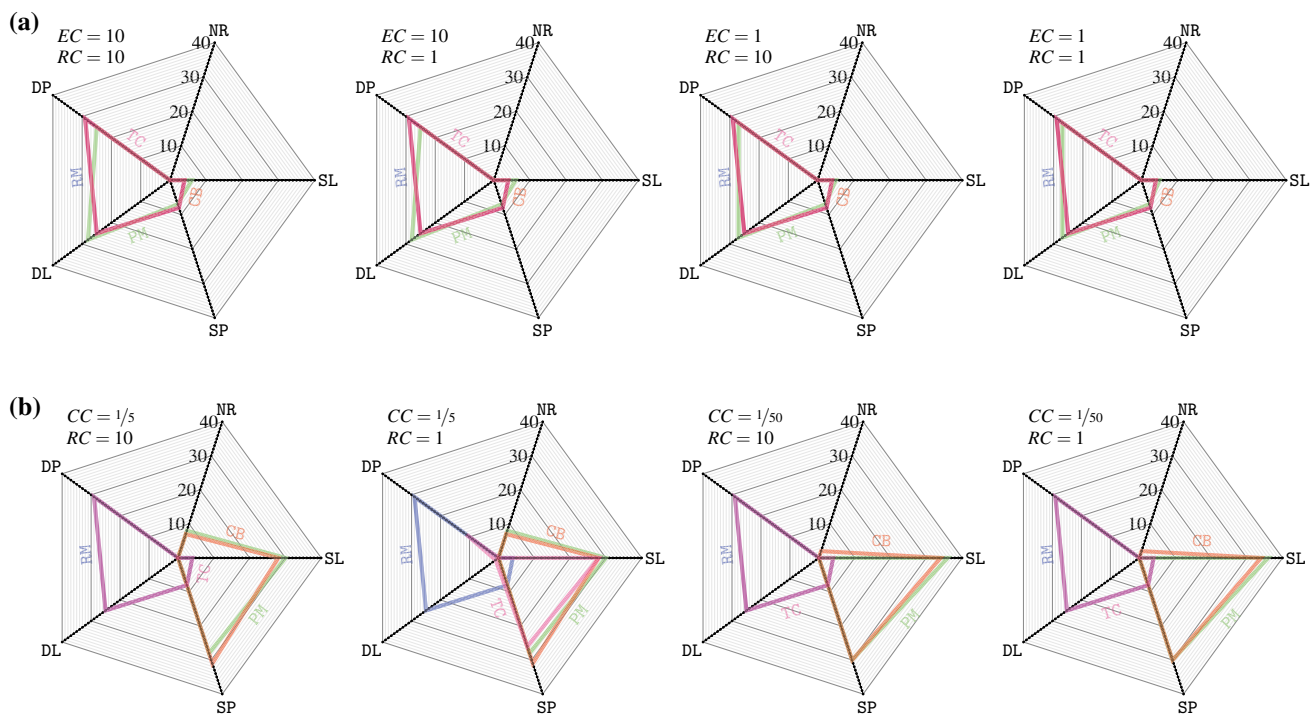


Fig. 7 Summary of results for various risk mitigation strategies found with the fine-grained optimization based on various monetary coefficients. The RA strategy always applies NR to all the 66 demands in the PL network. **a** Involved recovery budget is based on energy usage. EC

monetary energy coefficient. RC monetary risk coefficient. **b** Involved recovery budget is based on the amount of reserved capacity. CC monetary capacity coefficient. RC monetary risk coefficient

used. For comparison, the lower part relates to optimization when it is conducted on the basis of the reserved capacity. Note that in the case of risk minimization, the method of budget calculation (energy or capacity) and the monetary budget coefficient do not matter. There are three basic facts seen if we optimize on the basis of the actual energy used. (1) The optimal solutions for all the strategies and coefficient combinations are very similar (the distributions of recovery options do not change, and the assignment of different options to connections is steady). For various coefficients, the resulting (B, R) values differ, so that if we decrease the cost coefficient 10 times, then B decreases approximately 10 times, too. (2) The optimal solutions mainly combine two recovery options (both dedicated protections, excluding the shared link protection which seemed to be attractive due to the coarse-grained optimization; see Fig. 4). For a few demands, shared protection performs better than dedicated protection. The differences are very small and stem from more advantageous sharing of resources in backup links. This is again due to the fact the energy costs are not proportional to the capacities used. (3) The combination of the recovery options is also very similar across the coefficients representing monetary cost. Therefore, to find the fine-grained optimum, the best option is to look for profit maximization, which is almost identical to solutions to other strategies. It is based on the greatest ratio of the unit energy budget per unit decrease in

risk found among various recovery options for each connection separately. This approach follows theoretical results in the net utility maximization theory as mentioned in Sect. 4. In the case of fine-grained optimization based on capacity reservation, the dependence on the monetary coefficients is more important, and there is no easy way to find a single solution for all the risk mitigation strategies or risk measures applied (a fact reported also in [16]).

In Fig. 8, diagrams in the (B, R) space show selected results, again confirming the overshadowing options if we base the budget on energy cost (the effect is also shown with the coarse-grained approach described in Sect. 5). Again, this stems from the fact that additional energy usage over NR for providing recovery is low and similar across various recovery options due to the relatively low occurrence of failures when backup resources need to be used. This effect is additionally made stronger by the concavity of the energy profiles. The situation would be different if the budget were based on monetary costs proportional to reserved capacity, as typically done in network dimensioning. Related results are shown in Fig. 8 and in the lower part of Fig. 7, where more variability is seen for the capacity-based solutions. To make both situations comparable, the capacity and energy monetary coefficients are selected so that the total cost for non-recovered connections is of the same order as the cost when the budget is based on energy usage. Although risk

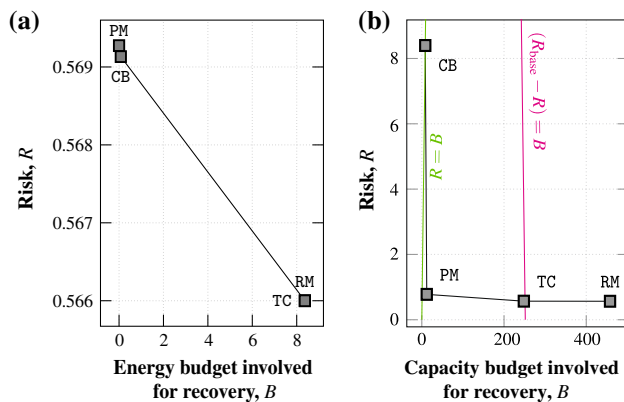


Fig. 8 Results presenting various response strategies with fine optimization. None of the optimal solutions are dominated by another. The energy and risk monetary coefficients equal 1. The capacity coefficient equals $1/50$. **a** Budget based on energy usage. **b** Budget based on capacity reservation

parameters do not change, the budget structure is quite different and changes the character of the results. We can see that the solutions differ, and even NR appears in the combinations responsible for risk mitigation. There is no simple method of using profit maximization as a universal basis for finding solution to risk mitigation. In Fig. 8b, it can even be seen that we are able to find solutions for cost balance and total benefit coverage by looking for non-dominated solutions and characteristic lines shown in Fig. 3. Therefore, when the budget is based on the costs proportional to reserved capacities, the theoretical characteristic assumed in Sect. 4 is not degenerated. However, as can be seen, the absence of degradation is in fact a negative effect from the practical perspective.

8 Conclusions and challenges

We present a framework for dealing with the assignment of recovery methods in energy-efficient resilient networks using the sleep mode. As energy-efficiency is meant to be traded off with resiliency, we propose to overtake this trade-off by using concepts elaborated in risk engineering. First, to find the optimal flows for recovery methods, we use the properties of typical energy profiles (concave and $f+p$) and then we are able to apply an effective Yaged's heuristic. Then, after expressing energy and risk with monetary equivalents, we show the properties of various solutions of respected risk mitigation strategies used in business management. Surprisingly, in the case of networks dealing with sleep modes, the solutions to these strategies are typically reduced to almost the same configurations, providing network designers with simple rules of thumb independently of the assumed compensation policies, risk measures, topologies assumed, and

monetary relationships. Since it is sometimes claimed that the sleep mode is not relevant for backup resources, such an option is also checked from the risk engineering viewpoint. It is shown that under some conditions, this is a feasible and advantageous option. We show that the intuitions related to the recovery costs based on capacity reservation are not necessarily valid when these costs are based on energy usage. For instance, as the lengths of paths are typically longer in the energy-efficient networks, the approach where shorter connections are recovered with re-routing, and longer ones with protections, is not necessarily useful in general. It is simply better to be directed with the type of reservation costs: when they are non-negligible, a more reasonable decision is to apply protections for all the connections. On the other hand, if the capacity reservation is free of charge or its cost is not significant, then it may be reasonable to concentrate working connections in a small number of links to be able to put others to sleep and save energy. This will cost a negligible increase in the total penalty risk.

We perceive the following problems as our further work: (a) we currently assume that in each scenario, all the demands apply the same compensation policy, although these differ in real situations; it is necessary to elaborate the modeling that allows us to calculate risk measures for other connections differentiated from this perspective; (b) the optimization model proposed for combining recovery options is naïve (although it appears to be useful); we predict that the iterative method (cyclically repeating the risk assessment and risk response stages as in the risk management cycle) that combines interchanging optimization and simulation with updates of the risk measures and the recovery assignment can be convergent and will provide exact risk and energy usage data; (c) here, there is only one budget dimension taken into account; however, it may be also important to take into account the combination of the monetary cost of energy usage and leasing the capacity, as it is crucial in the protection versus re-routing decision; (d) energy costs and profiles may change hourly or daily; multi-time-period design can be used to model more exactly the aim to provide energy-efficiency; (e) another option for improving the precision of the models presented is to add more exact energy profiles describing various elements in the optical network, also involving the opaque model, multiple layers, RWA constraints, or limited amount of capacity installed in the links.

Acknowledgments The authors would like to thank P. Boryło, P. Guzik, A. Kamisiński, A. Lasoń, K. Rusek, J. Rzasa, and A. Szymański for useful discussions related to the topic of this paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- [1] Ackermann, T.: IT Security Risk Management. Perceived IT Security Risks in the Context of Cloud Computing. Springer, Berlin (2013)
- [2] Aleksić, S.: Energy-efficient communication networks for improved global energy productivity. *Telecommun. Syst.* **54**(2), 183–199 (2013)
- [3] Anderson, E.E.: Firm objectives, IT alignment, and information security. *IBM J. Res. Dev.* **54**(3), 5 (2010)
- [4] Artzner, P., et al.: Coherent measures of risk. *Math. Financ.* **9**(3), 203–228 (1999)
- [5] Bao, N.H., et al.: Power-aware provisioning strategy with shared path protection in optical WDM networks. *Opt. Fiber Technol.* **18**(2), 81–87 (2012)
- [6] Barbaro, A., Bagajewicz, M.J.: Managing financial risk in planning under uncertainty. *AIChE J.* **50**(5), 963–989 (2004)
- [7] Bianzino, A.P., et al.: A survey of green networking research. *IEEE Commun. Surv. Tutor.* **14**(1), 3–20 (2012)
- [8] Bolla, R., et al.: Designing optimal energy profiles for network hardware. In: *Proceedings of GLOBECOM'12*. Anaheim, CA (2012)
- [9] Bolla, R., et al.: Energy efficiency in the future Internet: a survey of existing approaches and trends in energy-aware fixed network infrastructures. *IEEE Commun. Surv. Tutor.* **12**(2), 223–244 (2011)
- [10] Bolla, R., et al.: Enabling backbone networks to sleep. *IEEE Netw.* **25**(2), 26–31 (2011)
- [11] Cano, E.L., et al.: Energy efficiency and risk management in public buildings: strategic model for robust planning. *Comput. Manag. Sci.* **11**(1–2), 25–44 (2014)
- [12] Caria, M., et al.: A comparative performance study of load adaptive energy saving schemes for IP-over-WDM networks. *IEEE/OSA J. Opt. Commun. Netw.* **4**(3), 152–164 (2012)
- [13] Cavdar, C., et al.: Energy-efficient design of survivable WDM networks with shared backup. In: *Proceedings of GLOBECOM'10*. Miami, FL (2010)
- [14] Chiaraviglio, L., et al.: Modeling sleep mode gains in energy-aware networks. *Comput. Netw.* **57**(15), 3051–3066 (2013)
- [15] Cholda, P., et al.: Continuity-based resilient communication. In: *Proceedings DRCN 2009*. Washington, D.C. (2009)
- [16] Cholda, P., et al.: Risk mitigation in resilient networks. In: *Proceedings of RNDM 2014*. Barcelona, Spain (2014)
- [17] Cholda, P., et al.: Risk-awareness in resilient networks design: value-at-risk is enough. In: *Proceedings of NETWORKS 2014*. Funchal, Portugal (2014)
- [18] Cholda, P.: Risk-aware design and management of resilient networks. In: *Proceedings of RISI 2014*. Fribourg, Switzerland (2014)
- [19] Cholda, P., et al.: A survey of resilience differentiation frameworks in communication networks. *IEEE Commun. Surv. Tutor.* **9**(4), 32–55 (2007)
- [20] Cholda, P., Jajszczyk, A.: Recovery and its quality in multi-layer networks. *IEEE/OSA J. Light. Technol.* **28**(4), 372–389 (2010)
- [21] Cholda, P., et al.: Towards risk-aware communications networking. *Reliab. Eng. Syst. Saf.* **109**, 160–174 (2013)
- [22] Claise, B., et al.: Energy object context MIB. IETF Internet-draft (work in progress) (2014)
- [23] Clemente, R., et al.: Risk management in availability SLA. In: *Proceedings of DRCN 2005*. Lacco Ameno, Island of Ischia, Italy (2005)
- [24] Colman-Meixner, C., et al.: Disaster-survivable cloud-network mapping. *Photon Netw. Commun.* **27**(1), 141–153 (2014)
- [25] Dikbiyik, F., et al.: Minimizing the disaster risk in optical telecom networks. In: *Proceedings of OFC/NFOEC 2012*. Los Angeles, CA (2012)
- [26] Elnegeard, N.K., Stordahl, K.: Modelling uncertainty and risk in telecommunication investment project. *Elektronikk* **104**(3/4), 119–135 (2008)
- [27] Fan, M., et al.: Choosing a project risk-handling strategy: an analytical model. *Int. J. Prod. Econ.* **112**(2), 700–713 (2008)
- [28] Francini, A., Stiliadis, D.: Rate adaptation for energy-efficiency in packet networks. *Bell Labs Tech. J.* **15**(2), 131–146 (2010)
- [29] Francois, F., et al.: Leveraging MPLS backup paths for distributed energy-aware traffic engineering. *IEEE Trans. Netw. Serv. Manag.* **11**(2), 235–249 (2014)
- [30] Franke, U.: Optimal IT service availability: shorter outages, or fewer? *IEEE Trans. Netw. Serv. Manag.* **9**(1), 22–33 (2012)
- [31] Garvey, P.R.: *Analytical Methods for Risk Assessment. A Systems Engineering Perspective*. CRC Press, Boca Raton (2009)
- [32] Gonzalez, A.J., Helvik, B.E.: Guaranteeing SLA availability in telecommunications networks. In: *Proceedings of NETWORKS 2012*. Rome, Italy (2012)
- [33] Gonzalez, A.J., Helvik, B.E.: SLA success probability assessment in networks with correlated failures. *Comput. Comm.* **36**(6), 708–717 (2013)
- [34] Idzikowski, F., et al.: TREND in energy-aware adaptive routing solutions. *IEEE Commun. Mag.* **51**(11), 94–104 (2013)
- [35] IEEE Std 1490™-2003: IEEE guide adoption of PMI standard. A guide to the Project Management Body of Knowledge (2004)
- [36] Jirattigalachote, A., et al.: Dynamic provisioning strategies for energy efficient WDM networks with dedicated path protection. *Opt. Switch. Netw.* **8**(3), 201–203 (2011)
- [37] Kilper, D., et al.: Energy challenges in current and future optical transmission networks. *Proc. IEEE* **100**(5), 1168–1187 (2012)
- [38] Liu, Y., et al.: Delay-SRLG constrained, backup-shared path protection in WDM networks with sleep scheduling. *Comput. Comm.* **36**(2), 211–222 (2013)
- [39] López, J., et al.: Differentiated quality of protection to improve energy efficiency of survivable optical transport networks. In: *Proceedings of OFC/NFOEC 2013*. Anaheim, CA (2013)
- [40] Maillé, P., Tuffin, B.: *Telecommunication Network Economics. From Theory to Applications*. Cambridge University Press, Cambridge (2014)
- [41] Mansini, R., et al.: Twenty years of linear programming based portfolio optimization. *Eur. J. Oper. Res.* **234**(2), 518–535 (2014)
- [42] Markopoulou, A., et al.: Characterization of failures in an operational IP backbone network. *IEEE/ACM Trans. Netw.* **16**(4), 749–762 (2008)
- [43] Mastroeni, L., Naldi, M.: Compensation policies and risk in service level agreements: a value-at-risk approach under the ON-OFF service model. In: *Proceedings of ICQT 2011*. Paris, France (2011)
- [44] Morea, A., et al.: Protocol enhancements for “greening” optical networks. *Bell Labs Tech. J.* **18**(3), 211–230 (2013)
- [45] Muhammad, A., et al.: Energy-efficient WDM network planning with dedicated protection resources in sleep mode. In: *Proceedings of GLOBECOM'10*. Miami, FL (2010)
- [46] Musumeci, F., et al.: A power consumption analysis for IP-over-WDM core network architectures. *IEEE/OSA J. Opt. Commun. Netw.* **4**(2), 108–117 (2012)
- [47] Musumeci, F., et al.: Energy-efficiency of protected IP-over-WDM networks with sleep-mode devices. *J. High Speed Netw.* **19**(1), 19–32 (2013)
- [48] Ogino, N., Nakamura, H.: Telecommunications network planning method based on probabilistic risk assessment. *IEICE Trans. Commun.* **E94-B**(12), 3459–3470 (2011)

- [49] Orgerie, A.C., et al.: A survey on techniques for improving the energy efficiency of large-scale distributed systems. *ACM Comput. Surv.* **46**(4), 47:1–47:31 (2014)
- [50] Perelló, J., et al.: Link vs. opto-electronic device sleep mode approaches in survivable green optical networks. In: *Proceedings of OFC/NFOEC 2013*. Anaheim, CA (2013)
- [51] Pióro, M., Medhi, D.: *Routing, Flow and Capacity Design in Communication and Computer Networks*. Morgan Kaufmann Publishers, Los Altos (2004)
- [52] Quittek, J., et al.: Requirements for energy management. *IETF RFC 6988* (2013)
- [53] Restrepo, J.C.C., et al.: Energy profile aware routing. In: *Proceedings of GreenComm'09*. Dresden, Germany (2009)
- [54] Ricciardi, S., et al.: Modeling energy consumption in next-generation wireless access-over-WDM networks with hybrid power sources. *Math. Comput. Model.* **58**(5–6), 1389–1404 (2013)
- [55] Schupke, D.A.: Multilayer and multidomain resilience in optical networks. *Proc. IEEE* **100**(5), 1140–1148 (2012)
- [56] Shen, G., Tucker, R.S.: Energy-minimized design for IP over WDM networks. *IEEE/OSA J. Opt. Commun. Netw.* **1**(1), 176–186 (2009)
- [57] Swanson, M., et al.: Contingency planning guide for federal information systems. NIST SP 800–34 Rev. 1. National Institute of Standards and Technology, Gaithersburg (2010)
- [58] Vajanapoom, K., et al.: Risk based resilient network design. *Telecommun. Syst.* **52**(2), 799–811 (2013)
- [59] Van Heddeghem, W., et al.: Power consumption modeling in optical multilayer networks. *Photonic Netw. Commun.* **24**(2), 86–102 (2012)
- [60] Vasseur, J.P., et al.: *Network Recovery, Protection and Restoration of Optical, SONET-SDH, IP, and MPLS*. Morgan Kaufmann Publishers, Los Altos (2004)
- [61] Vizcaíno, J.L., et al.: Protection in optical transport networks with fixed and flexible grid: cost and energy efficiency evaluation. *Opt. Switch. Netw.* **11**(part B), 55–71 (2014)
- [62] Wiatr, P., et al.: Energy efficiency and reliability tradeoff in optical core networks. In: *Proceedings of OFC 2014*. San Francisco, CA (2014)
- [63] Wiatr, P., et al.: Green WDM-PONs: exploiting traffic diversity to guarantee packet delay limitation. In: *Proceedings of ONDM 2013*. Brest, France (2013)
- [64] Wiatr, P., et al.: Power savings versus network performance in dynamically provisioned WDM networks. *IEEE Commun. Mag.* **43**(5), 48–55 (2012)
- [65] Wickboldt, J.A., et al.: A framework for risk assessment based on analysis of historical information of workflow execution in IT systems. *Comput. Netw.* **55**(13), 2954–2975 (2011)
- [66] Xia, M., et al.: Risk-aware routing for optical transport networks. In: *Proceedings of INFOCOM 2010*. San Diego, CA (2010)
- [67] Xia, M., et al.: Green provisioning for optical WDM networks. *IEEE J. Sel. Top. Quantum Electron.* **17**(2), 437–445 (2011)
- [68] You, F., et al.: Risk management for a global supply chain planning under uncertainty: models and algorithms. *AIChE J.* **55**(4), 931–946 (2009)
- [69] Zhang, Y., Fan, Z.P.: An optimization method for selecting project risk response strategies. *Int. J. Qual. Proj. Manag.* **32**(3), 412–422 (2014)
- [70] Zhao, Y., et al.: Load balance vs. energy efficiency in traffic engineering: a game theoretical perspective. In: *Proceedings of INFOCOM 2013*. Turin, Italy (2013)



Piotr Cholda graduated from AGH University of Science and Technology (Kraków, Poland) as a master of science in networking in 2001. Afterward, he obtained a doctorate in telecommunications in 2006 from the same university. Then, he joined the Department of Telecommunications there, and now he works as an Assistant Professor. He specializes in design of computer and communications networks. Although his background

comes from technology, as he started his research in the reliability modeling and optimization of resilient multilayer long-haul networks (like the ones based on GMPLS, that is, generalized multiprotocol label switching), later he moved toward softer problems, like the effective provisioning of services to the network clients. Thus, he extended his interests to problems embracing the so-called quality of recovery (QoR), resilience-based differentiation in communications/computer networks, and QoS/QoR for peer-to-peer networking. Recently, Piotr has focused on risk-based communications networking, as he perceives this area as a very promising research field, where the interface between business needs and threats from one side and the resilience technology from the other can be used to obtain a practical equilibrium where the balance between client needs and technological possibilities can be found. He is the co-author of sixteen refereed journal papers and two conference tutorials. He is the recipient of the Communications QoS, Reliability and Performance Modeling Symposium Best Paper Award from ICC'06. He worked on network recovery problems in some of the 6. and 7. FP European Projects (IP Nobel/Nobel II, NoE EuroNGI/EuroFGI/EuroNF, SmoothIT). He was a Technical Program Committee (TPC) Co-Chair of NGI 2008 and NGI 2012, a TPC Co-Chair of Communications QoS, Reliability and Modelling Symposium at ICC 2011, a TPC Co-Chair of DRCN 2011, and a TPC Co-Chair of NOMS 2014. He is a member of the editorial board for IEEE Communications Surveys and Tutorials and leads the Book Review Column in IEEE Communications Magazine. He is a member of the ACM and IEEE.



Piotr Jaglarz was born in 1993 in Kraków. In 2012, he began to study Information and Communication Technologies at the Faculty of Computer Science, Electronics and Telecommunications (AGH University of Science and Technology, Kraków, Poland).